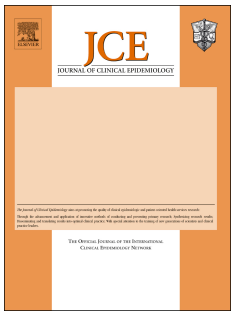
Accepted Manuscript



Quasi-experimental study designs series – Paper 4: uses and value

Till Bärnighausen, Peter Tugwell, John-Arne Røttingen, Ian Shemilt, Peter Rockers, Pascal Geldsetzer, John Lavis, Jeremy Grimshaw, Karen Daniels, Annette Brown, Jacob Bor, Jeffery Tanner, Arash Rashidian, Mauricio Barreto, Sebastian Vollmer, Rifat Atun

PII: S0895-4356(17)30283-4

DOI: 10.1016/j.jclinepi.2017.03.012

Reference: JCE 9348

To appear in: *Journal of Clinical Epidemiology*

Received Date: 29 September 2015

Revised Date: 18 March 2017

Accepted Date: 21 March 2017

Please cite this article as: Bärnighausen T, Tugwell P, Røttingen J-A, Shemilt I, Rockers P, Geldsetzer P, Lavis J, Grimshaw J, Daniels K, Brown A, Bor J, Tanner J, Rashidian A, Barreto M, Vollmer S, Atun R, Quasi-experimental study designs series – Paper 4: uses and value, *Journal of Clinical Epidemiology* (2017), doi: 10.1016/j.jclinepi.2017.03.012.

Quasi-experimental study designs series – Paper 4: uses and value

Till Bärnighausen[1,2,3], Peter Tugwell[4], John-Arne Røttingen[5], Ian Shemilt[6], Peter Rockers[7], Pascal Geldsetzer[2], John Lavis[2,8], Jeremy Grimshaw[9], Karen Daniels[10], Annette Brown[11], Jacob Bor[12], Jeffery Tanner[13], Arash Rashidian[14], Mauricio Barreto[15,16], Sebastian Vollmer[17], Rifat Atun[2]

[1]Institute of Public Health, Faculty of Medicine, Heidelberg University, Heidelberg German

[2]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, USA

[3]Africa Health Research Institute, Somkhele, South Africa

[4]Centre for Global Health, University of Ottawa, Canada

[5]Department of Health Management and Health Economics, University of Oslo, Oslo, Norway

[6]institute of Education, University College London, London, UK

[7]Center for Global Health and Development, Boston University, Boston, USA

[8]McMaster Health Forum, Centre for Health Economics and Policy Analysis, Department of Health Research Methods, Evidence and Impact, and Department of Political Science, McMaster University

[9]Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada

[10]Health Systems Research Unit, South African Medical Research Council, Tygerberg, South Africa

[11]Resaerch and Evaluation Strategic Initiative, FHI360, Washington DC, USA

[12]School of Public Health, Boston University, Boston, USA

[13]Impact Evaluation, World Bank, Washington DC, USA

[14]Department of Information, Evidence and Research, World Health Organization, Geneva, Switzerland

[15]Instituto de Saúde Coletiva, Federal University of Bahia, Salvador, Brazil

[16]National Institute in Science, Technology and Innovation in Health (CITECS), Salvador, Brazil

[17]Department of Economics, University of Göttingen, Göttingen, Germany

Abstract

Quasi-experimental studies are increasingly used to establish causal relationships in epidemiology and health systems research. Quasi-experimental studies offer important opportunities to increase and improve evidence on causal effects: (i) they can generate causal evidence when randomized controlled trials are impossible; (ii) they typically generate causal evidence with a high degree of external validity; (iii) they avoid the threats to internal validity that arise when participants in non-blinded experiments change their behavior in response to the experimental assignment to either intervention or control arm (such as compensatory rivalry or resentful demoralization); (iv) they are often well-suited to generate causal evidence on long-term health outcomes of an intervention, as well as non-health outcomes such as economic and social consequences; and (v) they can often generate evidence faster and at lower cost than experiments and other intervention studies.

What's new

- Quasi-experiments are often seen as a second-best design option when experiments are not feasible.

- Quasi-experiments are indeed often a good alternative to generating strong causal evidence, when experiments are not possible because of
    1. Ethical or political constraints to experimental assignment to intervention and control group
    2. Funding constraints
    3. Time constraints

- However, quasi-experiments also have distinct advantages over experiments that can render them the first-best option in certain instances. These advantages include the following:
    1. Quasi-experiments can generate causal evidence with higher external validity than experiments, because they typically distort the context of an intervention less than experiments.
    2. Quasi-experiments avoid threats to internal validity that can arise when participants in non-blinded experiments change their behavior in response to the experimental assignment (such as compensatory rivalry or resentful demoralization)

Introduction

Randomized controlled experiments are commonly considered the "gold standard" for causal evidence in health research [1]. Randomization of individuals to treatment and control groups ensures that treatment assignment is exogenous, i.e., no factor affecting the outcomes of interest can have exerted any influence on treatment assignment. As a result of randomization, selection bias in effect estimation is eliminated and the treatment and control group differ, in expectation, only in the treatment assignment. As corollary, all confounding factors, both observed and unobserved, are automatically balanced across treatment arms, and thus do not need to be separately controlled for in the estimation of causal effects of the treatment on outcomes.

Traditionally, for evidence synthesis in health, randomized controlled trials (RCTs) have been considered the main type of study that is of sufficient strength in generating causal evidence to warrant inclusion in systematic reviews and meta-analyses. For instance, the *Cochrane Collaboration* used to assert that "[r]andomized trials are the preferred design for studying the effects of healthcare interventions because, in most circumstances, the randomized trial is the study design that is least likely to be biased," while contending that exceptionally "review authors may be justified in including non-randomized studies," if "the question of interest cannot be answered by randomized trials" [2]. More recently, however, the *Cochrane Collaboration* has started an initiative on "assessing risk of bias in non-randomised studies" (ROBINS-I), which it plans to extend in the future to "specific types of NRSI [non-randomised studies of interventions], such as self-controlled designs, controlled before-and-after studies, interrupted time series studies, and studies based on regression discontinuity and instrumental variable analyses" [3]. The specific types of studies mentioned in this quote are sometimes referred to as quasi-experiments because – like experiments – they offer a strategy to circumvent the bias that can arise from participants' endogenous selection into different treatment arms, i.e., selection based on factors that also affect the outcomes of interest. If certain assumptions are met, quasi-experiments can generate evidence of causal strength similar to that of RCTs [4]. Empirically, comparisons of quasi-experiments and experiments have demonstrated that causal effect size estimates in these two types of studies are of similar size [5-8]. Quasi-experiments are thus a viable alternative for causal inference when RCTs are not possible. For this reason, there have been recent calls for including quasi-experiments in evidence synthesis [9-12], in particular in the field of health systems research where interventions, such as national health care reforms, are unlikely to be amenable to RCT-based evaluation.

However, quasi-experiments are more than just an alternative to consider when RCTs are not feasible. Because quasi-experiments combine some of the advantages of investigator-led and controlled trials with those of non-experimental research, they offer several distinct advantages over RCTs. In particular, they avoid some of the threats to external and internal validity that can arise in non-blinded RCTs. They are furthermore well-suited to generate causal evidence at low cost and on long-term outcomes. Below, we first define quasi-experiments for the purpose of this article and then discuss five scientific opportunities that quasi-experiments offer.

While quasi-experiments have several advantages over other study designs, they will only generate valid inferences when certain assumptions are met that are particular to the specific types of quasi-experiments. The assumptions for a number of types of important quasi-experiments are described in detail and critically discussed in another article in this

themed issue of the *Journal of Clinical Epidemiology*, on "Quasi-experimental study designs for evaluating practice, programs and policies: assessing the assumptions" [4]. Here, we will focus on the potential uses and values of quasi-experiments for causal impact evaluation of health practice, programs and policies. Other papers in this themed issue discuss why and how quasi-experimental results should be included alongside experimental results in evidence synthesis [13-18].

Defining quasi-experiments

There have been varying definitions of quasi-experiments in the intellectual history of the concept [19, 20]. For the purpose of this article, we have chosen the definition by King et al. of a quasi-experiment as "an observational study with an exogenous explanatory variable [or treatment, or exposure] that the investigator does not control" [21], which is consistent with other definitions published in the past two decades [22, 23]. Exogeneity of exposure means that the exposure is not influenced by the outcome of interest or any variable that is associated with the outcome. Exogeneity of exposure implies that selection bias and confounding is controlled for, without needing to observe and explicitly control for any confounding factors in the analysis.

In randomized controlled experiments, exogeneity is achieved through the investigator's randomization of individuals to treatment and control groups. In quasi-experiments, exogeneity is plausible for other reasons, including nature, policy, and practice. The credibility of a particular application of a type of quasi-experiment, such as instrumental variable or regression discontinuity designs, hinges on the plausibility of key assumptions. These assumptions are typically stronger than the assumptions necessary for an RCT to generate valid results, but they are also often weaker than the unconfoundedness assumption that needs to be met for non-experimental observational to generate unbiased inferences [24].

According to the above definition, the following study designs are quasi-experimental: instrumental variable designs (which identify exogenous variation in a treatment by using a variable that is associated with the exposure but is not independently associated with the outcome of interest) [25], r*egression discontinuity* designs (which can be used when a treatment is assigned according to a threshold rule and exploits exogenous exposure around the threshold value of an observed variable) [26, 27], and interrupted time series, as variant of regression discontinuity [28]. The definition excludes strategies to identify causal effects without exogenous treatment assignment, which cannot control for any unobserved confounding factors, such as regression, stratification, or matching studies – i.e., non-experiments [29].

Two types of boundary cases are included in the definition:

- Boundary cases between randomized controlled experiments and quasi-experiments are *policy experiments* which use randomization to allocate individuals to different "treatments", but for purposes other than establishing causal effects and without involving scientific investigators in the randomization. For instance, policies are sometimes assigned randomly because random assignment is considered a fair process, such as in the case of the US Vietnam draft lotter [30, 31].

- Boundary cases between quasi-experiments and non-experiments are *fixed effects* studies, including the special case of *difference-in-differences* (or controlled before-and-after) studies. These studies can control for some but not all unobserved confounding (e.g., unobserved time-invariant but not time-varying confounding [32] or unobserved sibling-invariant but not sibling varying confounding [33]), and they have thus been called "weak" quasi-experiments [29].

According to King's definition, quasi-experiments combine features of experiments (exogenous exposure) and non-experiments (observation without an investigator's intervention). In comparison to non-experiments, they thus have the potential to provide causal evidence of similar strength to that generated in RCTs; in contrast to RCTs, this evidence can be generated in an observational study. As "observational trials", quasi-experiments can add substantial value to health research and evidence synthesis. In particular, they can establish intervention effects in 'real life' – and thus with high external validity –, following discovery and RCT-based efficacy testing. And they can substitute for RCTs, when ethics, operational constraints or politics disallow active trial research. Below, we discuss five uses of quasi-experiments to illustrate their value as part of our methodological armamentarium, complementing both experiments and non-experiments in the quest to establish causal impacts of health practice, programs and policies.

Use 1: Generating causal evidence when randomized controlled trials are impossible

Quasi-experiments can often be carried out when evidence on size or magnitude of a causal effect are unknown and RCTs cannot be carried out [34]. This situation may arise for ethical, operational, or political reasons.

- *Ethical reasons*
  One requirement that typically needs to be met for an RCT to be considered ethically permissible is equipoise [35], despite criticism of this requirement [36]. Equipoise is genuine uncertainty as to whether one treatment is better than another one. In many cases, the directionality of a causal effect is known, eliminating equipoise, but the magnitude of the causal effect remains uncertain. The magnitude of causal effects is important, however, for individual decision-making weighing benefits of a treatment against potential harms and for policy-making allocating scarce resources across different treatments. Another example when RCTs will unethical, even though causal effect size estimates will have value, are trials of harmful exposures. Such trials will violate the ethical requirements of "do no harm" and equipoise.
- *Operational reasons*
  RCTs may not be practically feasible in many circumstances. In health systems research, major interventions commonly occur at the national level, such national health systems reforms [37] or health policy changes, such increases in tobacco taxation. While causal evidence of these national interventions would be highly valuable, it will typically be operationally impossible to randomly assign these national interventions to intervention and control groups, because no one can be practicably excluded from the intervention.
- *Political reasons*
  Even if an RCT is practically feasible an ethically sounds, it may be politically unwelcome. For instance, for the evaluation of the national Sure Start program in England, the investigators considered an RCT to have been the strongest approach

to causal evaluation but could not carry out such a trial because "government decisions precluded this possibility" [38].

In all three cases, quasi-experiments may provide opportunities to gain strong causal evidence when RCTs cannot be carried out. Quasi-experiments are ethically permissible to establish the magnitude of causal effect sizes if equipoise has been eliminated, because the investigator does not control treatment assignment. For instance, Bor et al. used a regression discontinuity approach to quantify for the first time the causal effect of immediate vs. deferred antiretroviral treatment on mortality in a sub-Saharan African community [27], after it had been firmly established that the treatment was a highly efficacious strategy for reducing HIV mortality. Quasi-experiments may also be feasible when randomization is not because an intervention cannot be withheld from anyone in country either because it is truly national in nature or because policy makers are opposed to random intervention allocation.

Use 2: Generating causal evidence with high degrees of external validity

Quasi-experiments typically use data on all patients or other "entire" populations across all available settings; the data for quasi-experiments is usually collected using routine data systems such as clinical records or population census data; and the intervention under study in quasi-experiments is normally delivered through real-life systems with routinely available resources and in every-day delivery contexts. In contrast, randomized controlled experiments select sites and patients and interfere in the intervention processes.

Quasi-experiments are thus more likely to generate causal evidence that applies to intervention implementation in real life. Indeed, a common complaint by clinicians regarding evidence from RCTs and evidence synthesis based only on RCTs is that it is unclear whether the evidence is relevant to their routine practice [39]. RCTs can lead to externally invalid results because of selection effects and "artificiality" introduced at different stages of the trial [40]. Selection effects arise because RCTs do not enrol populations that are representative of the populations that the trial results are intended to be generalized to. Artificiality arises because RCTs include procedures and processes that will not exist in the eventual real-life implementation of the RCT-tested intervention. Both selection effects and artificiality have several sources.

Sources of selection effects in RCTs

- Site selection: RCTs typically take place in a few selected study sites. An upper bound of the number of sites included in an RCT is the number of patients to be enrolled in the trial, which is constrained because of limited budgets and because the ethical obligation to minimize the risk of harm to trial participants implies that the number of participants should not exceed the minimum number required to detect significant intervention effects. The sample of sites participating in a trial is commonly not representative of all sites [41], because site staff need to agree to trial participation. In contrast, quasi-experiments are not constrained in sample size usually use routine data from all sites where an intervention is provided.
- Patient selection: RCTs typically apply inclusion and exclusion criteria in selecting trial participants, creating systematic differences between trial participants and the

population receiving an intervention in real life. Even if a trial does not apply any selection criteria, trial participants are likely to be a selected group, because trials rely on people volunteering to participate, and the volunteers may not be representative of the entire population who will receive an intervention in real life [42, 43]. In contrast, quasi-experiments normally do not require any selection criteria.

Sources of artificiality in trials

- Informed consent process: Participation in RCTs normally requires informed consent and the processes leading to informed consent may affect intervention effects. For instance, the informed consent process may put more emphasis on adverse events and side effects of an intervention than the information given about the intervention in routine delivery. Information and other aspects of informed consent processes may change the way that trial participants act. For instance, after the informed consent process trial participants may be more scared of potential side effects and adverse events than they would have been had they received the same intervention in routine delivery in the 'real world'. As a result, trial participants may be less likely to be adherent to the intervention than people receiving the intervention in the 'real world'. Quasi-experiments, by contrast, rely on routinely collected data and normally do not need to consent people for participation. As a result, they avoid this source of artificiality.
- Delivery processes: RCTs are commonly carried out by adding health workers and other resources to routine health systems to manage participant enrolment and to collect trial-specific data. Additionally, RCTs typically include other non-routine processes, such as frequent provider training and intensified adverse event monitoring and follow-up. These enhancements of delivery of care in both intervention and control arms of a trial render the trial context different from the 'real world' context in which an intervention will eventually be delivered (and in which quasi-experiments typically take place). These contextual changes threaten the external validity of trial results.

While selection effects and artificiality are likely to threaten external validity to a larger extent in experiments than in quasi-experiments, the results from some types of experiments will likely be more generalizable than the results from other types. For instance, pragmatic trials aim to minimize both selection effects and artificiality, by restricting the number of inclusion and exclusion criteria, selecting a wide range of sites for trial participation, and limiting interventions in the processes used to deliver an intervention [44]. But even pragmatic trials require some infrastructure and some processes that would not exist without the trial and may distort the context of intervention delivery. In contrast – because of their observational nature – quasi-experiments typically avoid both the selection effects and the artificiality of controlled experiments. One important use of quasi-experiments is thus 'real-life' effectiveness testing following intervention discovery and RCT-based efficacy testing [45]. In policy practice, proof of efficacy established in the artificial contexts of RCTs is often sufficient to lead to policy adoption of an intervention. In this case, quasi-experiments offer opportunities to confirm the intervention effectiveness in routine implementation.

Use 3: Avoiding threats to internal validity that can plague experiments

In non-blinded RCTs, internal validity in effect size estimation can be threatened, because trial participants react to their assignment to the treatment or the control group. Quasi-experiments are typically "blinded" and thus avoid such threats.

- *Compensatory rivalry:* Subjects in control arms of non-blinded RCTs sometimes behave abnormally because social competition motivates them to attempt to reduce anticipated treatment effects. Such so-called John Henry effects [46] are unlikely when subjects do not know that outcomes are observed and used to establish treatment effectiveness, as is generally the case in quasi-experiments.
- *Resentful demoralization:* Conversely, subjects in control arms of non-blinded RCTs may become resentful of not receiving the intervention the RCT aims to test. As a result, they may behave in ways that affects outcomes negatively [47], such as ceasing to carry out activities necessary for standard-of-care, leading to an upward bias in intervention effect size estimates. Again, in quasi-experimental studies, subjects are unlikely to know that outcomes are observed and that they have been exogenously "assigned" to treatment and control groups, reducing the likelihood of biases due to resentful demoralization.

In addition, it is sometimes possible to use more than one quasi-experimental approach in the same study [48]. Similar results across different quasi-experimental studies will strengthen our confidence in the internal validity of findings.

Use 4: Generating evidence on long-term and non-health outcomes

RCTs in health research commonly assess causal intervention effects over a time horizon of only a few years For instance, the vast majority of the more than 100 RCTs that have tested interventions to enhance adherence to HIV anti-retroviral treatment ran for five years or less and focused on clinical proxy indicators [49, 50]. However, often the effectiveness of an intervention over far longer time horizons is important, as in the case of HIV treatment, which is life-long. Quasi-experiments typically use routinely collected data and can thus provide evidence on causal effects over the long-term beyond typical RCT time horizons. Another advantage of basing causal evidence on data that is routinely and continuously collected is that causal effects can be repeatedly established as contexts change over time. In particular interventions whose effectiveness is likely to depend on contextual factors affected by beliefs, attitudes, and habits, such as behavioral interventions [51], may change in their (short-term) effects over time, and effect sizes established a long time ago may no longer be valid. In this situation, even if theory and non-experimental evidence suggest that intervention effects might have changed over time, given strong prior RCT evidence it is unlikely that equipoise can be established to ensure ethicality of future RCTs. Quasi-experiments are well-suited to repeatedly determine causal effects as the distance in time from initial RCT-based results increases. Finally, in health research, RCTs typically aim to establish causal effects on a few key health outcomes, but are rarely designed to determine 'broader' economic and social outcomes. For policy-making, both within the health sector and across sectors, evidence on economic and social causal impacts of health interventions is likely to be important [52]. Some interventions – such as the treatments of the chronic diseases of old age – may have little impact on a economic and social outcomes, while others – such as childhood vaccination – may affect outcomes across a person's entire life

course, including education, income and social functioning. Once an RCT has demonstrated primary causal effects of an intervention on health outcomes, future RCTs testing causal impact on economic and social outcomes will be unethical, because health interventions of proven effectiveness in improving health cannot be ethically withheld from study participants. In this case, quasi-experiments offer powerful opportunities to establish causal impacts beyond health [53].

Use 5: Generating evidence fast and at low cost

As discussed above, quasi-experiments offer several powerful opportunities to increase or improve causal evidence when RCTs are either not feasible or are not the right approach for generating a particular type of result. Even if RCTs are feasible and the right approach, however, quasi-experiments may have potential to contribute to the evidence base. Because quasi-experiments are typically retrospective and use routinely collected data, when they are possible, they can be a fast and low-cost approach to determine causal effects.

RCTs are clearly necessary in for efficacy in clinical medicine and, if feasible, should also be increasingly carried out to test health practice, programs and policies [54-56]. However, RCTs for the real-life causal impact evaluation are frequently expensive and generate results only after several years. Quasi-experiments thus offer opportunities for establishing causal evidence when such evidence is required quickly [57] or funding for an RCT is unlikely to be available.

Discussion

This paper is in its essence a discussion; however, three particular points warrant a separate discussion section: competing terminologies, the uptake of quasi-experiments and experiments in different fields of science, replication studies comparing causal estimates from randomized to quasi-experimental approaches, and the inclusion of quasi-experimental results in evidence synthesis.

Competing terminologies

While the term “quasi-experiments” has been commonly used in the way that we use it in this article – i.e., an observational study with an exogenous exposure that the investigator does not control” [21-23] – other uses of the term have also been common. Importantly, some authors have used the word “quasi-experiment” to describe different sets of approaches to causal inference. For instance, Greenberg and Schroder call cluster randomized controlled trials “quasi-experiments” [58], and Shadish, Cook and Campbell define “quasi-experiments” to include all designs that have a “comparison group”, whether assignment to intervention and comparison group is exogenous or not [59]. On the other hand, some authors, such as Dunning, use the term “natural experiments” to describe designs we have called “strong quasi-experiments”: “policy experiments”, regression discontinuity designs, and instrumental-variable designs [60]. Given the multitude of definitions, researchers using terms such as “quasi-experiments” or “natural experiments” need to carefully describe what the mean when using them. While the multitude of definitions

and the fact that they typically describe partially overlapping sets of designs can be an obstacle to debate and consensus, fundamentally it is not important. What is important is the clarity of basic conceptual distinctions in causal inference, in particular, the conceptual distinctions between designs that can control for all, some, or no unmeasured confounders.

Quasi-experiments and experiments

Quasi-experimental studies have been widely accepted and used in the social sciences for several decades [61, 62]. Only more recently, they are also gaining increasing traction in health research, in large part because of the opportunities to generate novel and better evidence that we discuss above [27, 63-65]. In this context, it is interesting to note that while there is a movement to increasingly use quasi-experiments in health research, in the social sciences experiments are increasingly becoming a methodological standard. In neither of these broad scientific fields, however, has the adoption of a class of research methods into the standard methodological tool set meant a departure from the other class. The social sciences are still generating numerous and important research findings using quasi-experiments, while increasingly also carrying out experiments; the health sciences are unlikely to decrease the use of experiments in establishing treatment efficacy of novel technologies as the increasingly realize quasi-experimental opportunities for causal inference. Rather, the adoption of a new methods class opens up opportunities to answer research questions that could not be answered with the previously used class of methods. For instance, in the health sciences, an important driver of the use of quasi-experimental methods has been the "impact evaluation" agenda [66, 67], which focuses on how to establish the "real-life" impact of medical technologies whose efficacy has been established in randomized controlled experiments [68, 69].

An increasing number of studies have compared causal effect size estimates yielded by randomized trials to those of non-experimental and quasi-experimental approaches. While a comprehensive review of this literature here is beyond the purview of this article, this literature suggests that regression discontinuity designs perform well in direct comparisons to randomized trials [5-7]. The evidence comparing other types of quasi-experimental and non-experimental approaches to randomized trials is more mixed [8, 70], and will depend on the degree to which the assumptions of each method are met [71].

Population-based experiments

At the same time, in the social sciences so-called population-based experiments have been commonly used. Such experiments aim to overcome the external validity limitations of "facility-based" experiments by enrolling random samples of well-defined populations in the experiment, e.g., all adults in a nation. Population-based experiments have been particularly commonly used in population-based survey experiments, in which different questions are randomly allocated to respondents as random stimuli [72]. For instance, randomly allocation of alternative approaches to frame a question can be used to establish whether the age, sex or race of a person displaying a certain behavior matters for the respondent's evaluation of that behavior. However, theoretically many types of stimuli other than survey questions can be randomly allocated to entire populations. Population-based experiments are likely to eliminate some of the threats to external validity that "facility-based" experiments commonly suffer from. In particular, some of the selection processes reducing external validity are eliminated when a population-representative sample is visited in their homes and asked to

participate in an experiment. Other threats to external validity, however, may persist even if a representative sample of an underlying well-defined population participates in an experiment. In particular, experiments – whether population- or facility-based – do require an intervention that would be absent in "real life". The intervention and its delivery mechanism (e.g., a survey) can change the context in which the intervention effect is measured, reducing the strength of our belief that the intervention will have the same effect in "real life" that was established in the experiment.

Encouragement experiments

One of the major arguments for the increasing attention and use of quasi-experiments – the fact that experiments are not ethically permissible when equipoise is violated – can sometimes be overcome through the use of so-called encouragement experiments [73, 74]. In these studies, a treatment, which is known to be beneficial, is randomly encouraged but participants are explicitly allowed to decide whether to receive the treatment or not. Examples of encouragements that are randomly allocated in these experiments include intensified counselling on the benefits of a treatment, over and above counselling in the standard of care, or a financial incentive to take up the treatment. The example of encouragement experiments demonstrates again that the definitions of quasi-experiments and experiments are fluid and that it is useful to distinguish between identification and estimation strategy. While encouragement experiments are legitimately called experiments (because a researcher randomly assigns an exposure), the random exposure assignment serves to generate an instrumental variable, i.e., the encouragement can be used as an instrument to estimate the causal effect of the encouraged treatment on an outcome of interest in an instrumental variable approach.

Quasi-experiments in evidence synthesis

The arguments for quasi-experimental studies we outline above are arguments for both primary quasi-experimental research and better integration of quasi-experimental results in evidence synthesis. In particular, the potential quasi-experiments have to provide strong causal evidence when RCTs are not possible; the strength of quasi-experiments in producing externally valid results; and the potential quasi-experiments have to establish causal effects on long-term and non-health outcomes should be considered when debating whether to include quasi-experimental results in the synthesis of a body of evidence. Current practices synthesizing bodies of evidence commonly ignore quasi-experimental results. Guidelines for evidence synthesis – e.g. the GRADE criteria (http://www.gradeworkinggroup.org) – and for reporting of studies – e.g., STROBE guidelines – do not make separate recommendations for quasi-experimental studies. Quasi-experiments are typically graded as observational, non-experimental studies, but this assessment does not allow that some quasi-experiments, e.g. regression discontinuity designs and randomized natural or policy experiments, yield inferences nearly as strong as randomized clinical trials. Furthermore, the features of a rigorously-presented quasi-experimental study differ from the features of a rigorously-presented cohort or case-control study. Devising checklists for the reporting and assessment of quasi-experiments is difficult due to the wide range of data generating processes underlying quasi-experiments and the wide range of methods used to analyze them [75, 76]. Often, specific subject area knowledge is required to determine the plausibility of assumptions invoked in a quasi-experiment. And yet, the potential of these designs to produce rigorous, actionable evidence

to improve patient care, policy design, and resource allocation is too great to be ignored.

One important area for future research is how to best integrate quasi-experiments into methods for evidence synthesis [11, 29]. Stronger integration of quasi-experimental results into evidence synthesis may have powerful behavioral feedback effects. The organizations that establish standards and guidelines for evidence synthesis, such as the Cochrane Collaboration, send signals to primary researchers about what is considered evidence of sufficiently high quality and causal strength to influence policy and health care practice. Increasing acceptance and consideration of quasi-experiments as a legitimate source of causal results for evidence synthesis in health is thus likely to increase the supply of quasi-experimental evidence. Quasi-experiments have great potential to generate novel and important insights, but it is likely that they are currently substantially underutilized relative to this potential. Increasing use of quasi-experimental results in evidence synthesis may contribute to closing the gap between the potential and the realized contributions of quasi-experimental studies in informing health care practice, programs and policy.

Conclusion

Quasi-experiments offer the practical advantages that they can be carried out when randomized experiments are not possible. They have the important advantages that they typically generate results that are of higher external validity than experimental results, because they take place in ‘real world’ settings rather than in the artificial context of experiments. They further avoid the threats to internal validity that arise when participants in non-blinded experiments change their behavior in response to the experimental assignment, such as compensatory rivalry or resentful demoralization. Quasi-experiments are also well suited to establish causal effects on long-term health outcomes, as well as on non-health outcomes of a health intervention, such as social and economic consequences. Finally, quasi-experiments often generate results faster and at lower costs than experiments. It is likely that quasi-experiments will be increasingly used in epidemiology and health systems research. Quasi-experimental results should thus be considered and integrated in systematic reviews, meta-analyses and other evidence syntheses of causal effects of health care practice, programs and policy.

References

1. Kaptchuk, T.J., *The double-blind, randomized, placebo-controlled trial: gold standard or golden calf?* J Clin Epidemiol, 2001. **54**(6): p. 541-9.
2. Higgins, J. and S. Green, *Cochrane handbook for systematic reviews of interventions*. Vol. Version 5.1. 2011: The Cochrane Collaboration.
3. Sterne, J.A., et al., *ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions.* BMJ, 2016. **355**: p. i4919.
4. Bärnighausen, T., et al., *Quasi-experimental study designs for evaluating practice, programs and policies: assessing the assumptions.* Journal of Clinical Epidemiology (this issue), 2017.
5. Aiken, L.S., et al., *Comparison of a Randomized and Two Quasi-Experimental Designs in a Single Outcome Evaluation: Efficacy of a University-Level Remedial Writing Program.* Evaluation Review, 1998. **22**(2): p. 207-244.
6. Green, D.P., et al., *Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks.* Political Analysis, 2009. **17**(4): p. 400-417.
7. Cook, T.D., W.R. Shadish, and V.C. Wong, *Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons.* Journal of Policy Analysis and Management, 2008. **27**(4): p. 724-750.
8. Fortson, K., et al., *Using an Experimental Evaluation of Charter Schools to Test Whether Nonexperimental Comparison Group Methods Can Replicate Experimental Impact Estimates*. 2012, Mathematica Policy Research Institute: Washington D.C.
9. Rockers, P.C. and T. Bärnighausen, *Interventions for hiring, retaining and training district health systems managers in low- and middle-income countries.* Cochrane Database Syst Rev, 2013. **4**: p. CD009035.
10. Rockers, P.C., et al., *Study-design selection criteria in systematic reviews of effectiveness of health systems interventions and reforms: A meta-review.* Health Policy, 2012. **104**(3): p. 206-14.
11. Langlois, E.V., et al., *Advancing the field of health systems research synthesis.* Syst Rev, 2015. **4**: p. 90.
12. Tugwell, P., et al., *Assessing equity in systematic reviews: realising the recommendations of the Commission on Social Determinants of Health.* BMJ, 2010. **341**: p. c4739.
13. Aloe, A.M., et al., *Quasi-experimental study designs for evaluating practice, programs and policies: Collecting data from quasi-experimental studies.* Journal of Clinical Epidemiology (this issue), 2017.
14. Becker, B.J., et al., *Quasi-experimental study designs for evaluating practice, programs and policies: Synthesizing evidence for effects collected from quasi-experimental studies presents surmountable challenges.* Journal of Clinical epidemiology (this issue), 2017.
15. Glanville, J., et al., *Quasi-experimental study designs for evaluating practice, programs and policies: Identifying quasi-experimental studies to inform systematic reviews.* Journal of Clinical Epidemiology (this issue), 2017.
16. Lavis, J.N., T. Bärnighausen, and F. El-Jardali, *Quasi-experimental study designs for evaluating practice, programs and policies: Supporting the production and use of health systems research syntheses that*

draw on quasi-experimental study designs. Journal of Clinical Epidemiology (this issue), 2017.
17. Rockers, P.C., et al., *Quasi-experimental study designs for evaluating practice, programs and policies: Strengthening global capacity for evidence synthesis of quasi-experimental health systems research.* Journal of Clinical Epidemiology (this issue), 2017.

18. Rockers, P.C., et al., *Quasi-Experimental Study Designs for Evaluating Practice, Programs and Policies: Realizing the Full Potential of Quasi-Experiments for Health Research.* Journal of Clinical Epidemiology (this issue), 2017.
19. Campbell, D.T. and J.C. Stanley, *Experimental and quasi-experimental designs for research*. 1996, Boston: Houghton Mifflin Company.
20. Shadish, W.R., T.D. Cook, and D.T. Campbell, *Experimental and quasiexperimental designs for generalized causal inference*. 2002, Boston: Houghton Miffin.
21. King, G., R.O. Keohane, and S. Verba, *The importance of research design in political science.* American Political Science Review, 1995. **89**(2): p. 475-481.
22. Meyer, B.D., *Natural and quasi-experiments in economics.* Journal of Business & Economic Statistics, 1995. **13**(2): p. 151-161.
23. Remler, D. and G. Van Ryzin, *Research methods in practice: strategies for description and causation*. 2010, Thousand Oaks, CA: Sage Publications.
24. Bärnighausen, T., et al., *Quasi-experimental study designs for evaluating practice, programs and policies: assessing the assumptions.* Journal of Clinical Epidemiology, 2017.
25. Angrist, J.D., G.W. Imbens, and D.B. Rubin, *Identification of causal effects using instrumental variables.* Journal of the American Statistical Association, 1996. **91**(434): p. 444-455.
26. Moscoe, E., J. Bor, and T. Bärnighausen, *Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice.* J Clin Epidemiol, 2015. **68**(2): p. 122-33.
27. Bor, J., et al., *Regression discontinuity designs in epidemiology: causal inference without randomized trials.* Epidemiology, 2014. **25**(5): p. 729-37.
28. Zuckerman, I.H., et al., *Application of regression-discontinuity analysis in pharmaceutical health services research.* Health Serv Res, 2006. **41**(2): p. 550-63.
29. Rockers, P.C., et al., *Inclusion of quasi-experimental studies in systematic reviews of health systems research.* Health Policy, 2015. **119**(4): p. 511-21.
30. Hearst, N., T.B. Newman, and S.B. Hulley, *Delayed effects of the military draft on mortality. A randomized natural experiment.* N Engl J Med, 1986. **314**(10): p. 620-4.
31. Hearst, N., et al., *The draft lottery and AIDS: evidence against increased intravenous drug use by Vietnam-era veterans.* Am J Epidemiol, 1991. **134**(5): p. 522-5.
32. Gunasekara, F.I., et al., *Fixed effects analysis of repeated measures data.* Int J Epidemiol, 2014. **43**(1): p. 264-9.
33. Anekwe, T.D., et al., *The effect of childhood measles vaccination on educational attainment: A mother-fixed-effects study in rural South Africa.* Vaccine, 2015.
34. Black, N., *Why we need observational studies to evaluate the effectiveness of health care.* BMJ, 1996. **312**(7040): p. 1215-1218.
35. Freedman, B., *Equipoise and the ethics of clinical research.* N Engl J Med, 1987. **317**(3): p. 141-5.
36. Miller, F.G. and H. Brody, *A critique of clinical equipoise. Therapeutic misconception in the ethics of clinical trials.* Hastings Cent Rep, 2003. **33**(3): p. 19-28.
37. Gottret, P., G. Schieber, and H. Waters, *Good practice in health financing : lessons from reforms in low and middle-income countries*. 2008, Washington, DC: World Bank.
38. Melhuish, E., et al., *Effects of fully-established Sure Start Local Programmes on 3-year-old children and their families living in England: a quasi-experimental observational study.* Lancet, 2008. **372**(9650): p. 1641-7.
39. Rothwell, P.M., *External validity of randomised controlled trials: "to whom do the results of this trial apply?".* Lancet, 2005. **365**(9453): p. 82-93.
40. Morton, R.B. and K.C. Willams, *Experimental political science and the study of causality: from nature to the lab*. 2010, New York: Cambridge University Press.
41. Rothwell, P.M., *Factors that can affect the external validity of randomised controlled trials.* PLoS Clin Trials, 2006. **1**(1): p. e9.

42. Melberg, H.O. and K. Humphreys, *Ineligibility and refusal to participate in randomised trials of treatments for drug dependence.* Drug Alcohol Rev, 2010. **29**(2): p. 193-201.
43. Fortin, M., et al., *Randomized controlled trials: do they have external validity for patients with multiple comorbidities?* Ann Fam Med, 2006. **4**(2): p. 104-8.
44. Tunis, S.R., D.B. Stryer, and C.M. Clancy, *Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy.* JAMA, 2003. **290**(12): p. 1624-32.
45. Abdulkadiroğlu, A., et al., *Accountability and Flexibility in Public Schools: Evidence from Boston's Charters And Pilots.* The Quarterly Journal of Economics, 2011. **126**(2): p. 699-748.
46. Saretsky, G., *The OEO P.C. experiment and the John Henry effect.* Phi Delta Kappan, 1972. **53**: p. 579–581.
47. Cook, T.D. and D.T. Campbell, *Quasi-experimentation: design and analysis issues for field settings*. 1979, Chicago: Rand-McNally.
48. Duflo, E., *Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment.* The American Economic Review, 2001. **91**(4): p. 795-813.
49. Chaiyachati, K.H., et al., *Interventions to improve adherence to antiretroviral therapy: a rapid systematic review.* AIDS, 2014. **28 Suppl 2**: p. S187-204.
50. Barnighausen, T., et al., *Interventions to increase antiretroviral adherence in sub-Saharan Africa: a systematic review of evaluation studies.* Lancet Infect Dis, 2011. **11**(12): p. 942-51.
51. Fishbein, M. and M. Yzer, *Using theory to design effective health behavior interventions.* Communication Theory, 2003. **13**(2): p. 164-183.
52. Barnighausen, T., et al., *Valuing vaccination.* Proc Natl Acad Sci U S A, 2014. **111**(34): p. 12313-9.
53. Anekwe, T.D., et al., *The causal effect of childhood measles vaccination on educational attainment: A mother fixed-effects study in rural South Africa.* Vaccine, 2015.
54. Finkelstein, A. and S. Taubman, *Health care policy. Randomize evaluations to improve health care delivery.* Science, 2015. **347**(6223): p. 720-2.
55. Tollefson, J., *Can randomized trials eliminate global poverty?* Nature, 2015. **524**(7564): p. 150-3.
56. King, G., et al., *A "politically robust" experimental design for public policy evaluation, with application to the Mexican Universal Health Insurance program.* Journal of Policy Analysis and Management, 2007. **26**(3): p. 479–506.
57. Adebamowo, C., et al., *Randomised controlled trials for Ebola: practical and ethical issues.* Lancet, 2014. **384**(9952): p. 1423-4.
58. Greenberg, D. and M. Schroder, *Digest of social experiments*. 1997, Washington, DC: Urban Institute Press.
59. Shadish, W., T. Cook, and D. Campbell, *Experimental and quasi-experimental designs for generalized causal inference*. 2002, Boston, MA: Hougthon Mifflin.
60. Dunning, T., *Natural experiments in the social sciences*. 2012, Cambridge, UK: Cambridge University Press.
61. Meyer, B., *Natural and quasi-experiments in economics.* J Bus Econ Stat., 1995. **13**: p. 151-161.
62. King, G., R.O. Keohane, and S.T. Verba, *The importance of research design in political science.* American Political Science Review, 1995. **89**: p. 454–481.
63. Hernan, M.A. and J.M. Robins, *Instruments for causal inference: an epidemiologist's dream?* Epidemiology, 2006. **17**(4): p. 360-72.
64. Tchetgen Tchetgen, E.J., et al., *Instrumental variable estimation in a survival context.* Epidemiology, 2015. **26**(3): p. 402-10.
65. Bor, J., et al., *Quasi-experiments to establish causal effects of HIV care and treatment and to improve the cascade of care.* Curr Opin HIV AIDS, 2015.

66. USAID, *Evaluation: learning from experience -- USAID evaluation policy*. 2011, Washington, DC: USAID.
67. Savedoff, W., R. Levine, and N. Birdsall, *When will we ever learn? Improving lives through impact evaluation*. 2006, Washington, DC: Center for Global Development.
68. Toulany, A., et al., *Quasi-Experimental designs for quality improvement research.* Implementation Science, 2012. **8**(Supplement 1): p. S3.
69. Kruk, M.E., *More health for the money–toward a more rigorous implementation science.* Sci Transl Med, 2014. **6**(245): p. 245ed17.
70. Glazerman, S., D.M. Levy, and D. Myers, *Nonexperimental versus Experimental Estimates of Earnings Impacts.* The Annals of the American Academy of Political and Social Science, 2003. **589**: p. 63-93.
71. Cook, T.D., P.M. Steiner, and S. Pohl, *How Bias Reduction Is Affected by Covariate Choice, Unreliability, and Mode of Data Analysis: Results From Two Types of Within-Study Comparisons.* Multivariate Behavioral Research, 2009. **44**(6): p. 828-847.
72. Mutz, D., *Population-based survey experiments*. 2011, Princetion, NJ: Princeton University Press.
73. Bradlow, E., *Encouragement designs: an approach to self-selected samples in an experimental design.* Marketing Letters, 1998. **9**(4): p. 383-391.
74. West, S.G., et al., *Alternatives to the randomized controlled trial.* Am J Public Health, 2008. **98**(8): p. 1359-66.
75. Duvendack, M., et al., *Assessing 'what works' in international development: meta-analysis for sophisticated dummies.* Journal of Development Effectiveness, 2012. **4**(3): p. 456-471.
76. Valentine, J. and H. Cooper, *A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: the study design and implementation assessment device.* Psychological Methods, 2008. **13**(2): p. 130–149.